ELSEVIER
Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Engineering 15 (2011) 1792 – 1796
Procedia Engineering
www.elsevier.com/locate/procedia

Advanced in Control Engineering and Information Science

Knowledge extraction from aerodynamic simulation data of compressor rotor

Wei Wang*, Rong Mo,Qingming Fan

The key laboratory of Contemporary Design and Integrated Manufacturing Technology, Northwestern Polytechnical University, Xi'an, 710072, China

Abstract

In order to acquire the knowledge which was embedded in the simulation data of the aerodynamic optimization of the compressor, the method based on the entropy measurement and decision tree algorithm was applied to NASA rotor37. The data attributes were reduced through calculating the entropy of the attributes of the experiment data set, and the rules of the optimization design were acquired by constructing the decision tree. The results indicate that the method of knowledge extraction is feasible in the filed of the optimization design of the compressor rotor.

Selection and/or peer-review under responsibility of [CEIS 2011]
Keywords: Knowledge extraction;Decision tree;Aerodynamic optimization;Simulation data;NASA rotor37;

1. Introduction

The optimization methods such as response surface methodology, genetic algorithm and simulated annealing are widely applied in the field of the aerodynamic optimization of the compressor rotor [1]. During the optimization process, a large number of simulation data are generated and massive implicit design knowledge is embedded in the simulation data. Accordingly, it is significant to extract the design knowledge from the simulation data through the knowledge discovery in database (KDD), for which can guide the optimization further and enhance the design efficiency [2].

In recent years, the methods of KDD and data mining were applied to the simulation data, e.g. extracting the knowledge from the simulated data of the driving head of hydraulic drill by the rough set

* * Corresponding author. Tel.:+86-15802956472;
E-mail address: yingccnp@163.com.

doi:10.1016/j.proeng.2011.08.333

theory (RST), proposing a novel method of knowledge acquisition in the field of plastic forming and obtaining the significance design rules [3,4]. In this research, knowledge discovery based on the entropy measurement and decision tree algorithm was applied to the simulation data of the compressor rotor. Through analysing the relationship between the aerodynamic performance of the compressor rotor and the geometric design parameters, and establishing the decision tree, the design rules of the compressor rotor were acquired. The method was universal that could be applied to other fields.

2. Framework of knowledge extraction from the optimization design simulation data of compressor rotor

According to the characteristics of the simulation data of compressor rotor, the framework of knowledge extraction was shown in Fig.1, including of the following main parts:

1. *The aerodynamic optimization of the compressor rotor.* First, the parametric designing model of the compressor rotor was built by the Autoblade software. Then the aerodynamic performance was calculated by the computational fluid dynamics (CFD) program. To study the relation between the design parameters and the aerodynamic performance of the compressor rotor, Design of Experiment (DOE) technology was applied.
2. *Simulation data collection.* During the iterative process of the blade optimization design, a large number of simulation data were generated, which could be used as the source for knowledge extraction. Therefore, the simulation data should be collected to build the simulation data database.
3. *Knowledge extraction.* The collected simulation data were incomplete, noisy and inconsistent. Moreover, the single data mining algorithm could not deal with all of the simulation data types. To improve the quality of the simulation data and satisfy the need of DM algorithm, data pre-processing must be made. For example, if a data mining (DM) algorithm could only handle the nominal attributes while the data sets such as the simulation data were real value, the discretization is often used as a pre-processing method to turn the simulation data into nominal data. At the stage of rule induction, data mining algorithms were applied in order to extract the knowledge. The common data-mining methods were the statistical methods, e.g. decision tress, artificial neural networks, and fuzzy set theory.
4. *Knowledge reuse.* The knowledge could help the designers understand simulation results more clearly, and be used as heuristic knowledge in the optimization process, and also could be used as a knowledge auto-acquisition tool to help knowledge designers in building knowledge base of a KBE system.

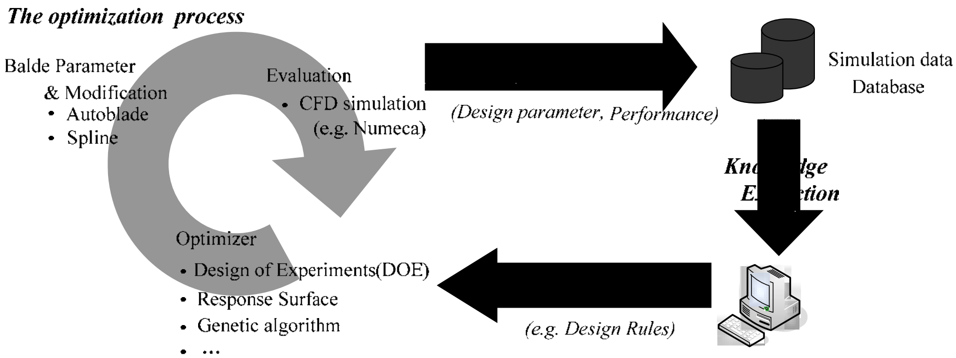

Fig.1 the framework of knowledge extraction from the simulation data of the compressor rotor

3. Knowledge extraction based on the entropy measurement and decision tree algorithm

3.1. Data attributes reduced base on the entropy measurement

According to the reduced algorithm of data attributes, the importance of data attributes were sequenced by the entropy measurement [5], the following steps were carried out:

Step a Creating the initial attribute data set: $F = \{f_i: \ f_i \in F, i = 1,2,...,n\}$

Step b Getting rid of the attribute f_i from the attribute data set F, and the relevant data set was acquired. Moreover, the formula of calculating the difference entropy between the attribute data set F and various subsets was:

$$E(s) = -\sum_{i}^{N-1}\sum_{j=i+1}^{N}\left\{S_{ij}\log_2 S_{ij} + \left(1-S_{ij}\right)\log_2\left(1-S_{ij}\right)\right\}$$

$$\textit{Where}: \quad S_{ij} = \begin{cases} \exp\left[\left(\ln 0.5/D\right)\left|X_i - X_j\right|\right] & \text{all of attributes are numerical} \\ \left(\sum_{k=1}^{n}\left|X_{ik} - X_{jk}\right|\right)\Big/ n & \text{others} \end{cases}$$

S_{ij} was the similar measurement between the sample X_i and X_j, N was the number of the sample data, n was the number of the data attributes.

Step c Selecting the attribute f_k which the difference entropy was minimal between the data set F and the subset F_{fi}, the new attribute set was acquired

Step d Repeating the process from the step a to c until just remaining one data attribute.

3.2. C5.0 decision tree algorithm

The decision tree algorithm classified all sample data according to the condition attributes. It began from the root node of the decision tree. Every attribute would generate two branches, the corresponding date sets belong to the scope of the branch attribute value will be removed from the new generated child node. The process would be iterated until all date sets both were classified. Every pathway from the root to the leaf node indicated one designing rule. If the decision tree was excessively complicated, we could use the pruning method to simplify the decision tree.

The C5.0 decision tree algorithm improved based on ID3, it classified the data set by the attribute of the node whose value of the information gain was maximum.

4. Application to NASA rotor 37 simulation data

4.1. NASA rotor 37 aerodynamic optimization

NASA rotor37 was the representative transonic compressor rotor which was designed by NASA Lewis research centre [6]. In this research, we fitted the gravity stack curve with the B-spine curve, and selected the coordinates of the B-spine control point in the sweep and leap plane as the design variable, see Fig.2, the pressure ratio as the objective. The Latin-optimal design of experiment method was applied to the optimization design of NASA rotor37, we did the 64 experiments in all. The difference values of the design parameters and compressor ratio between the experiment data and the initial design of NASA rotor37 were shown in Table 1.

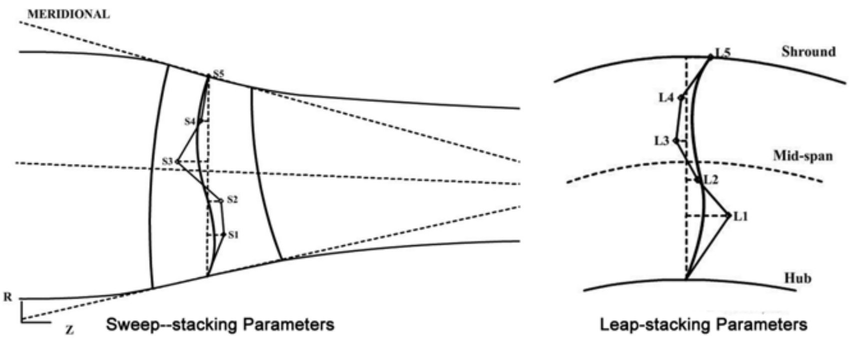

Fig.2. The design parameters in the sweep and leap planes

Table.1. The simulation data by the Latin-optimal design of experiment

No.	S1	S2	S3	S4	S5	H1	H2	H3	H4	H5	Pressure ratio
1	-0.0242	-0.3225	0.2282	-0.0974	-0.2648	0.0278	-0.0657	0.0597	0.0420	0.0096	>0
2	0.1374	-0.0558	0.2125	-0.0440	-0.3370	0.0585	0.0554	-0.0370	-0.0045	-0.0415	<0
3	0.1311	-0.3005	-0.1232	-0.1884	-0.1080	-0.0438	-0.0362	0.0515	-0.0616	-0.0390	>0
….	……					……					…
64	0.1939	-0.1217	-0.1577	-0.0064	-0.0013	0.0767	-0.0212	0.0138	-0.0308	-0.0454	<0

4.2. Knowledge extraction from the simulation data

The simulation data of the experiment design were discretized according to the cluster analysis and the attribute were reduced based on the entropy measurement. Then, we acquired the reduction of the attribute data sets, $F_f = \{S4, H5, H1, H2\}$, Whose information gain were 0.268,0.249,0.14,and 0.099 respectively.

According the decision tree algorithm, the root node was the attribute S4 whose information gain was maximum, the same way, the child nodes were the attributes H5, H1, and H2 respectively. At last the decision tree was built, see Fig.3.

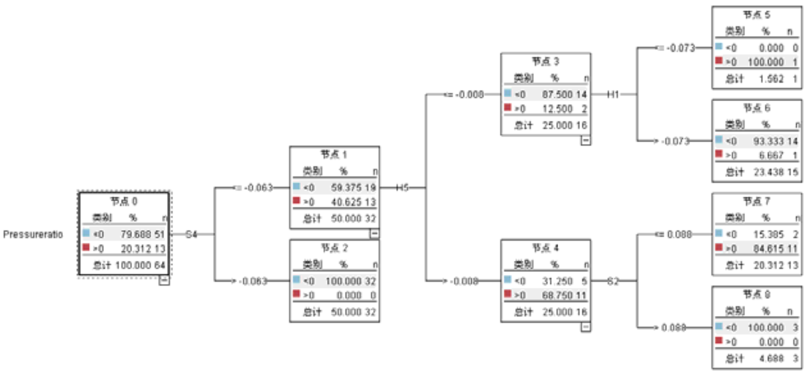

Fig.3. The decision tree of the compressor rotor

The structure of the decision tree was concise, so it needed not be pruned. According to the decision tree, the design rules were extracted which were shown in Table2.

Table.2. The design rules of extraction from the simulation data

No	the design rules	reliability
1	If S4>-0.063 Then Pressure ratio < 0	100%
2	If S4<=-0.063 and H5<=-0.0080 and H1<=-0.0732 Then Pressure ratio >0	100%
3	If S4<=-0.063 and H5<=-0.0080 and H1>-0.0732 Then Pressure ratio<0	93.33%
4	If S4<=-0.063 and H5>-0.0080 and S2>0.0885 Then Pressure ratio<0	100%
5	If S4<=-0.063 and H5>-0.0080 and S2<=0.0885 Then Pressure ratio>0	84.62%

The design rules indicated that the compressor rotor with the sweep and leap skewed had the significant influence on the aerodynamic performance, e.g. the pressure ratio would be improving by increasing the sweep-back (S4<=-0.063, S2<=0.0885) and leap-front (H5>-0.0080) of the compressor rotor.

5. Conclude

In this paper, a Framework of knowledge extraction from the optimization simulation data was established according to the characteristics of the simulation data. Then, the algorithm based on the entropy measurement and decision tree algorithm was applied to apply to the optimization simulation data of the compress rotor. Meanwhile, the decision tree was built, and the design rules were extraction concisely and refinedly which could be used as a guide for the further optimization of the compress rotor. In a word, the method of knowledge discovery from the simulation data was feasible in the field of the compressor rotor.

Acknowledgements

The study was supported by the National High-Tech. R&D Program (863 Program), China (No. 2007AA04Z184).

References

[1]. Chen L G, Sun F R. Optimum design of a subsonic axial-flow compressor stage. *Apllied Energy* 2005;80(2):187-195.
[2]. Hasenjäger,L.G.S.M. Knowledge Extraction from Knowledge Extraction from Aerodynamic Design Data and its Application to 3D Turbine Blade Geometries. *Journal of Mathematical Modelling and Algorithms* 2008;7:329-350.
[3]. Si-cong Yuan,Liu dao-hua,Huang Jun. Knowlege's acquisition method based on simulated data for driving head of hydraulic drill[J]. *Application Reasearch of Computers* 2009,26(5):1828-1831.
[4]. Yin Ji Long,Liu a Yong,Peng Ying hong. Fuzzy-rough Method of Knowledege Discovery on Numerical Simulation Results[J]. *Journal of Shang Hai Jiao Tong University* 2004;38(9):1448-1452.
[5]. Wang Wei,Yin Guo Fu, Long Jian Zhong. Application of data ming to cathode structure design by using finite element analysis. *Journal of Huangzhong University of Science and Technology(naturel science edition)* 2006;34(7):76-78.
[6]. Ahn C, K. Kim.Aerodynamic Design Optimization of an Axial Flow Compressor Rotor. *ASME Conference Proceedings* 2002. Amsterdam: 813-819.